Vyjádření školitele k disertační práci Jiřího Vábka

*Binární znaménkové reprezentace celých čísel v kryptoanalýze hashovacích funkcí*

Jiří Vábek zahájil své doktorské studium v akademickém roce 2005/06. Krátce po zahájení studia čínská badatelka Wang našla kolize v tehdy široce používané hashovaní funkci MD5. Jak je v kryptologii zvykem, výsledek pouze oznámila, nezveřejnila ale metodu, jak jej získala. Po nějaké době se podařilo z oznámených dvojic kolizních zpráv rekonstruovat diferenční cestu, která je klíčem pro algoritmus hledající kolizní zprávy. Tento algoritmus se různým badatelům podařilo urychlit a hledat kolize na PC během několika vteřin. Všechny tyto kolizní dvojice zpráv ale sledovaly stejnou diferenční cestu, kterou odhalila Wang. To znamená, že obě zprávy měly vždy délku dvou bloků a lišily se v šesti bitech (po třech v každém bloku) na pevně daných místech.

Různí autoři pak navrhovali jiné možnosti pro rozdíly v kolidujících zprávách, nikomu se ale nepodařilo skutečnou dvojici kolizních zpráv s těmito rozdíly najít.

O něco později Marc Stevens publikoval se spoluautory různé aplikace kolizí v MD5 založené na pravděpodobnostním algoritmu, který navrhnul. Na základě zveřejněného slovního popisu tohoto algoritmu se podařilo autorovi práce navrhnout vlastní verzi Stevensova algoritmu a spolu s Milanem Boháčkem ji úspěšně implementovali. Touto implementací se pak podařilo najít v té době zcela nový typ kolizních zpráv. Tento nový typ kolize autor disertace prezentoval na Konferenci INDOCRYPT 2008, příspěvek publikovaný v LCNS tvoří náplň třetí kapitoly práce.

Pro úspěšné použití algoritmu bylo důležité nastavit vhodně volitelné parametry. K tomu bylo nutné najít dobrý horní odhad počtu binárních znaménkových reprezentací celých čísel s danou nadváhou. Tento odhad pak umožnil vhodné nastavení volitelných parametrů tak, aby implementace algoritmu našla diferenční cestu a dvojici kolizních zpráv v přijatelném čase několika hodin na PC.

Tyto horní odhady jsou obsahem čtvrté kapitoly. Jak je z doby publikace algoritmu zřejmé, odhady byly použité už v při přípravě příspěvku na konferenci v roce 2008. Sepsání těchto odhadů trvalo několik let. Dílem proto, že Jiří Vábek začal pracovat na plný úvazek, dílem proto, že původní důkazy odhadů byly založené na extenzivním rozebírání jednotlivých případů. Důkaz založený na překladačích byl nakonec sepsán ve spěchu těsně před odevzdáním práce a bude muset před konečnou publikací ještě projít určitou revizí.

Poslední pátá kapitola obsahuje krátký vedlejší produkt práce na horních odhadech počtu binárních znaménkových reprezentací a ukazuje, že podobný odhad založený na jednoduché rekursivní posloupnosti platí i pro jiný číselný systém. Naznačuje-li tento výsledek skutečně možnost obecnějšího tvrzení pokrývajícího více číselných systémů ukáže až budoucnost.

Práce je uvedena stručným úvodem, který měl uvést čtenáře do problematiky hledání kolizí v hashovací funkci MD5 a propojit jednotlivé části práce. Částečně vychází z přehledového článku o kolizích v MD5 předneseného na mezinárodním workshopu Mikulášská kryptobesídka pořádaném v Praze v roce 2008 a publikovaném v jeho sborníku. Zde je třeba konstatovat, že tento záměr se úplně nezdařil a pro čtenáře, který není dobře obeznámen s konstrukcí hashovací funkce MD5 a problematikou hledání kolizí v ní, je hodně obtížně čitelný.

Jiří Tůma 15.8.2014